

TAGUNGSBERICHTE

den Zeitschriften sind vor allem Dissertationen und Nachschlagewerke zunehmend nur noch über das Internet verfügbar.

« »

**Jahrestagung der Gesellschaft für
Wissenschafts- und Technikforschung
Merton heute**

Wissenschaftsinterne Leistungskriterien, Evaluation und wissenschaftliche Praxis

Berlin, 1.- 2. Dezember 2006

**Tagungsbericht von Stefan Böschen,
Universität Augsburg**

Evaluation ist ‚in‘. Evaluation ist notwendig. Evaluation ist umstritten. Die Gründe für diese Situation sind vielschichtig. Zwei Aspekte sind sicherlich bemerkenswert. *Zum einen* sind Qualitätsurteile essenzieller Bestandteil der Produktion wissenschaftlichen Wissens, denn sie markieren das verlässliche Wissen, das Ausgangspunkt für die weitere Wissensproduktion ist. Die universalistische, uninteressierte und skeptische, also den Merton'schen Normen folgende Zuschreibung von Qualität bildet(e) zudem eine wesentliche Legitimationsgrundlage für die Sonderstellung von Wissenschaft in modernen Gesellschaften. *Zum anderen* wird der Wissenschaft in jüngster Zeit die Kontrolle über die Legitimation durch Qualität immer mehr entzogen und die internen Bewertungspraktiken durch eine externe Bewertung der Qualität wissenschaftlicher Arbeit ergänzt. Im Sog forschungspolitischer Schwerpunktsetzungen muss die Wissenschaft nachweisen, dass sie Qualität liefert und dies auch nach außen hin darzustellen vermag. Evaluationen werden so zu einem neuen Transmissionsriemen zwischen Wissenschaft und Gesellschaft. Doch welche Folgen hat dies für das Gefüge differenzierter wissenschaftlicher Erkenntnismärkte und das Verhältnis von Wissenschaft und Gesellschaft? Die Jahrestagung der Gesellschaft für Wissenschafts- und Technikforschung 2006

versuchte dieses Problemfeld zu konturieren und erste Antworten zu sammeln.

**1 Wandel von Lehre und Forschung durch
Evaluationsinstrumente**

Universitäten als Orte der Bildung und Forschung stehen in besonderer Weise im Mittelpunkt des Evaluationsgeschehens. Bildung wie Forschung sollen sehr gut, besser möglichst exzellent sein. Die Idee ist ja auch bestechend einfach: Warum sollten nicht die immer schon über das Peer Review vorgenommene Prüfungs- und Selektionsmechanismen innerhalb der Community ausgeweitet und dadurch eine gezielte Förderung der besten Forschungseinheiten möglich gemacht werden? In der Praxis des Universitätsalltags ist demgegenüber nicht nur die konkrete Gestaltung von Instrumenten der Evaluation alles andere als selbstverständlich, sondern es zeigen sich auch unerwartete Effekte und Nebenfolgen bei der Einführung und mehr oder minder forcierten Nutzung von evaluationsbasierten Verteilungsinstrumenten in den Hochschulen. Diese erste Sondierungsrunde der Tagung wurde dadurch strukturiert, dass anhand unterschiedlicher Dimensionen Form und Effekte von Evaluationsinstrumenten diskutiert wurden. Diese waren insbesondere

- die Varianten von Evaluationsformeln und der zeitliche Implementationsfortschritt
- der Einsatz von Evaluationsinstrumenten in verschiedenen Forschungsbereichen sowie
- bildungs- und forschungspolitische Konsequenzen

Letztere wurden einerseits allgemein im Rahmen einer Podiumsdiskussion diskutiert, andererseits konkret an dem neu gegründeten Institut für Forschungsinformation und Qualitätssicherung (IFQ, Vortrag Stefan Hornbostel) festgemacht.

1.1 Der zeitliche Implementationsfortschritt von evaluationsbasierten Instrumenten

Australien ist ein Land, in dem die Finanzierung der Universitäten in der Zwischenzeit schon wesentlich über Evaluation gesteuert

wird. Funktionsweise und mögliche Effekte und Nebenfolgen lassen sich hier, gleichsam wie in einem historischen Laboratorium, schon studieren. Entsprechend stellten sich Grit Laudel und Jochen Gläser (beide von der Australian National University, Canberra) das Thema „*Ein formelbasiertes Evaluationssystem – wie beeinflusst es die Forschungsinhalte? Die Finanzierung australischer Universitäten*“. Bei australischen Universitäten werden 24 % der Grundfinanzierung leistungsbasiert vergeben, wobei die stärkste Gewichtung der Indikatoren dem Drittmiteileinkommen zukommt. Vor dem Hintergrund einer Fülle von empirischen Arbeiten zeigten Laudel und Gläser auf, dass sich ein deutliches Anpassungsverhalten von Universitäten wie auch von Wissenschaftlern nachweisen lässt. Universitäten kopieren die staatliche Evaluations-Formel, investieren strategisch in die Drittmiteleinwerbung oder auch in die Bildung von so genannten kritischen Massen und bemessen die individuellen Forschungsleistungen entsprechend. Wissenschaftler verändern darauf hin ihre Arbeit, in dem sie ihr Forschungshandeln nach den Indikatoren orientieren, Forschungen aufgeben, die keine Drittmittelgeber fördern, oder Projekte ‚strecken‘. Dieser Effekt ist in weniger kostenintensiven Forschungsgebieten geringer ausgeprägt. Die Wissensproduktion schließlich orientiert sich stärker an ‚mainstream‘-Themen und -Anwendungskontexten. In der Summe zeigt sich die besonders exponierte Steuerungswirkung der Drittmittellandschaft, die zu einer Konzentration der Mittel auf wenige Universitäten und wenige Wissenschaftler beiträgt, und die starke Steuerungswirkung der Universitäten durch interne Leistungsbewertung und Zentrenbildung.

Einen anderen Eindruck über Forschen und Lehren an einer Hochschule erhielt man durch Stefan Langes (FernUniversität Hagen) Referat über eine Untersuchung darüber, wie wohl Forschung betrieben wurde (und heute noch wird), bevor Evaluationen eine Rolle gespielt haben: „*Stunde 0: Forschungsbedingungen und Zukunftserwartungen von Wissenschaftlern an einer deutschen Traditionsuniversität ohne kohärente Forschungsevaluation*“.¹ In dem untersuchten Beispiel wurden nur 20 % der Grundfinanzierung auf der Basis einer Performanz-Formel vergeben, die aber sehr stark Lehre und Größe gewichtete

(75 %). Aufgegliedert nach unterschiedlichen Fachbereichen (Wirtschaft, Naturwissenschaft, Philosophie) zeigte sich eine unterschiedliche Bereitschaft, mit evaluationsbasierten Instrumenten zu operieren, die vom erst- zum letztgenannten Fachbereich hin abnimmt. So zeigten sich auch hier schon erste Anzeichen eines „prä-evaluativen“ Stresses, insofern wachsende Zeitknappheiten über Drittmiteleinwerbung entstehen oder Profilbildung und ihre Nebenfolgen sichtbar werden. Darauf reagierten die befragten Wissenschaftler mit einer zeitlichen Dehnung von Forschung, ihrer Maßstabsverkleinerung oder mit einer Art „inneren Emigration“ – jedoch noch nicht mit einer Anpassung an die Indikatoren wie etwa des Publikationsverhaltens. In der Summe: „Merton ‚lebt‘ zwar noch an der deutschen Universität der vor-evaluativen Epoche, gerät aber partiell unter Druck.“

Gerd Grözing (Universität Flensburg) wandte sich, nachdem in den anderen beiden Vorträgen die in der allgemeinen Diskussion bekannten Qualitätsmaße wie Publikationen und Drittmittel zur Sprache gekommen waren, der Frage „*Zweitrufe als Attraktionsmaß?*“ zu. Zweitrufe sind alle weiteren Rufe an eine Hochschule, die eine bereits als Professor/in an einer Hochschule in Deutschland beruflich tätige Person in einem bestimmten Zeitraum erteilt werden. Die empirische Erhebung bestätigte u. a. die Erwartungen, dass Universitäten relativ mehr Zweitrufe als Fachhochschulen aufweisen und dass es einen (wenn auch schwachen) positiven Zusammenhang mit anderen Reputationsmessungen einer (öffentlichen) Hochschule gibt.

1.2 Kontexte der Forschung und Evaluation

Im Mittelpunkt der Überlegungen von Falk Schützenmeister (Technische Universität Dresden) zu „*Orientierung und Qualitätssicherung in der deutschen Universitätsforschung*“ stand die Frage nach dem Verhältnis von Wissenschafts- und Evaluationsforschung. Mittels einer Online-Umfrage unter deutschen Hochschullehrern/innen verdeutlichte er, dass Erhebung und Evaluation unterschiedlichen ‚Welten‘ angehören, wobei Eva-

uation als administrativer Versuch angesehen werden müsse, das Nichtsteuerbare zu steuern. Die (empirische) Wissenschaftsforschung hingegen versuche, Indikatoren für die Erklärung von Sachverhalten (etwa wissenschaftliche Produktivität) zu entwickeln. Dabei diskutierte er dies für die empirisch gewonnene Beobachtung, dass es zu einer Ausdifferenzierung zwischen drei Gruppen komme: der Gruppe der „Problemlöser“ (Wissenschaftler, bei denen die Lösung gesellschaftlicher Probleme dem Erkenntnisgewinn vorgezogen wird), der Gruppe der „Produktorientierten“ (Wissenschaftler, denen die Erkenntnis an erster Stelle steht, dann aber die wissenschaftliche Anwendbarkeit der Ergebnisse wichtiger ist als der gesellschaftliche Nutzen) und der Gruppe der „Mertonianer“ (Wissenschaftler, denen Erkenntnis und gesellschaftlicher Nutzen wichtiger sind als die wirtschaftliche Verwertung).

Jan-Hendrik Passoth (Universität Bielefeld) nahm die Anregung zu einer Auseinandersetzung mit Merton in theoretischer Weise auf. Sein Vortrag „*Und er wird die Fülle haben ...*“ – *Forschungsrankings und der Matthäus-Effekt*“ war insbesondere ein Plädoyer, sich nicht von zu einfachen Unterscheidungen in der Gegenüberstellung etwa von wissenschaftsinternen und wissenschaftsexternen Kriterien leiten zu lassen. Denn schon Merton habe sehr deutlich auf die vielschichtigen und durchaus auch kontraproduktiven Effekte wissenschaftlicher Systeme der Anerkennung und Zuweisung von Reputation hingewiesen. Seiner Auffassung nach ist die wesentliche Frage diejenige, ob in Bewertungen (wie etwa das Ranking des Centrums für Hochschulentwicklung (CHE) oder das Forschungsrating des Wissenschaftsrates) die Probleme, die auf einer Durchdringung der beiden Kriterienbereiche (intern – extern) beruhen, überhaupt reflexiv erfasst werden.

Die Differenzierung von Leistungskriterien gewinnt mit weiterer Anwendungsnähe an Brisanz. Dies war zumindest eine der Botschaften in den Ausführungen von Alexandra Manzai (Technische Universität Berlin) über „*APACHE-Score, DRG-System und Evidence Based Medicine. Probleme der Standardisierung und Technisierung von Wissen in einem Feld wissenschaftlicher Praxis: der Intensivmedizin*“. Ausgehend von der Beobachtung, dass die

Intensivmedizin durch ein vielschichtiges informationstechnologisches Netzwerk geprägt ist, diskutierte sie die Konsequenzen mit Blick auf die Standardisierung etwa durch spezifische medizinische Klassifikationssysteme (Scores). Die Pointe besteht darin, dass solche Systeme nicht einfach einen fokussierten Gegenstand (etwa Herz-Kreislauf-Funktionalitäten) abbilden, sondern eine spezifische Wirklichkeit erst herstellen. Vor diesem Hintergrund müssten eigentlich mittels ethnographischer Studien die Wirkungsweisen von Evaluationen empirisch untersucht werden – im Grunde sei eine Evaluation von Evaluationen (mittels immanenter Kriterien) notwendig.

1.3 Bildungs- und forschungspolitische Konsequenzen

Zunächst wurden im Rahmen einer Podiumsdiskussion (Sonja Berghoff, CHE; Jürgen Güdler, DFG, Johann Köppel, TU Berlin; Jochen Gläser, Australian National University of Canberra) der generelle Fragenkomplex „*Forschungsrankings: Notwendig? Unvermeidlich? Gut?*“ aufgeworfen. Dabei wurde ausgehend von der Beobachtung, dass Forschungsrankings gleichsam als Wachstumsindustrie betrachtet werden müssen, die Fragen diskutiert, wer denn überhaupt Rankings brauche, wie sie im Detail methodisch funktionieren, welche Auswirkungen von Rankings erhofft bzw. befürchtet werden und wie schließlich ein „ideales Ranking“ aussehen würde. Freilich war gerade die letzte Frage sehr pointiert, öffnete aber umso mehr den Blick auf die notwendige reflexive Einbettung von Rankings. Als forschungspolitisches Generalkonzept würden sie zu starken Verzerrungen durch Anpassungsverhalten führen, zugleich eröffnen Rankings steuerungsrelevante Einblicke in den ‚Forschungsbetrieb‘, die notwendige hochschulpolitische Qualitätsbemessungen und Konzentrationsbewegungen anleiten könnten. Wie herausfordernd im Einzelnen eine solche Aufgabe ist, verdeutlichte Stefan Hornbostel (IFQ) mit der Vorstellung seiner Einrichtung unter dem Titel „*Nähe und Distanz, Monitoring und Evaluation: Das Institut für Forschungsinformation und Qualitätssicherung*“. Dieses Institut, das auf eine Initiative der DFG zurückgeht, um

ein kontinuierliches Monitoring der Programmentwicklung sowie der Wirkungen des Förderhandelns zu ermöglichen, nahm im Oktober 2005 seine Arbeit auf. Im Wesentlichen verfolgt es drei Ziele:

- die Durchführung eines Förder- und Forschungsmonitorings, um eine dauerhafte Beobachtung von Entwicklungen im Sektor (öffentlich geförderter) Forschung anhand von validen Instrumenten (Kennzahlen und Indikatoren) zu ermöglichen;
- die Qualitätssicherung, um Effizienz und Effektivität von Förderprogrammen zu erfassen, aber auch Desiderata des Förderangebotes auszukundschaften;
- die Forschungsinformation, um einen Überblick über Akteure, Projekte und Leistungen der Forschung in Deutschland zu erhalten.

Die Gestaltung dieses anspruchsvollen Projektes ist allerdings nicht ganz einfach, sei es aus datenschutzrechtlichen Fragen beim Rückgriff auf die Antragsdatenbank der DFG, sei es aufgrund der Messproblematik von Evaluationen. Vor diesem Hintergrund führt das IFQ eigene Studien durch, wie z. B. eine Langzeitstudie zur Erfassung der Karrieren von Promovierenden in Abhängigkeit von unterschiedlichen Promotionsbedingungen; es entwickelt ein Forschungsinformationssystem (FINSYS) zur Dokumentation von Forschungsergebnissen aus DFG-geförderten Projekten, und nimmt aber auch selbst „Ad-hoc-Evaluationen“ wie die des Deutsch-Chinesischen Zentrums vor. Deutlich wurde, dass nur durch die Nutzung eines großen Spektrums von Methoden und Maßnahmen Evaluation in einem anspruchsvollen Sinne vorangebracht werden können.

2 Evaluation und transdisziplinäre Forschung

Einen weiteren Schwerpunkt bildeten Beiträge, die sich dem Thema der Evaluation transdisziplinärer Forschungsprozesse widmeten. Die Qualitätsmessung transdisziplinärer Forschung stellt ein ganz grundsätzliches Problem dar. Denn als Forschungsarbeit liegt sie definitionsgemäß jenseits der einzelnen Disziplinen. Welche Standards sollen also gelten? Die Beiträge der Tagung versuchten nicht nur

zu einer näheren Beschreibung des Problems zu gelangen, sondern auch konkrete Lösungsvorschläge zu unterbreiten. Den Auftakt machten Antonietta Di Giulio und Rico Defila (beide Universität Bern) mit ihren Überlegungen zu „*Inter- und transdisziplinäre Forschung evaluieren – Balance zwischen Leistungsmessung und Qualitätsmanagement*“. Sie rückten die Frage in den Mittelpunkt, was Evaluation mit Blick auf die Spezifika transdisziplinärer Forschung, die eben nicht allein verschiedene Disziplinen sondern darüber hinaus ja auch noch eine substantielle Beteiligung von Anwendern/innen vorsehen würde, überhaupt heißen könnte. Die zentrale Idee geht dahin, dass in der Evaluation nicht nur die Ergebnisse sondern ebenso der Prozess der Forschung mit zu berücksichtigen sind. Sowohl für Leistungsmessung als auch für Qualitätsmanagement gilt, dass die Beurteilung zwar nach Maßgabe wissenschaftlicher Kriterien erfolgt, diese aber projektspezifisch geleistet werden müsste. Transdisziplinäre Projekte müssten letztlich zu Beginn des Prozesses die Kriterien festlegen, anhand derer sie sich im Nachhinein auch messen lassen wollen.

In die konkrete Evaluationspraxis führten Alexander Walter, Sebastian Helgenberger, Arnim Wiek und Roland Scholz (alle von der ETH Zürich) mit ihrem Beitrag: „*Social impact evaluation of transdisciplinary research*“ ein. Vor dem Hintergrund des an der ETH Zürich praktizierten Modells transdisziplinärer Case Studies stellten sie eine Studie zur Evaluation sozialer Effekte eines solchen Forschungsprozesses vor. Dabei konnten sie zeigen, dass das Wissen hinsichtlich der mit dem Forschungsprozess verknüpften Entscheidungen von der Partizipationsintensität abhängig ist. Dieser Effekt erklärt sich vor allem über die Variablen ‚Netzwerkbildung‘ und ‚Transformationswissen‘, die zwischen Prozess und Entscheidungen vermitteln. Entsprechend lassen sich, so die generalisierende Schlussfolgerung, neben den wissenschaftlichen auch die sozialen Auswirkungen messen, jedoch müsste dies stärker prozessbegleitend geschehen und die entsprechenden Indikatoren validiert werden.

Stärker auf die Organisationsform ging Michael Guggenheim in seinem Beitrag „*Auf den Schultern von Experten: die normative Struktur außeruniversitärer Forschung und das*

Problem ihrer Bewertung“ ein. Am Beispiel von Firmen, die Umweltdienstleistungen anbieten, untermauerte er die These, dass die Verbindung von außeruniversitärer Organisationsform mit lokalen Forschungsgegenständen zur Etablierung von neuen Qualitätsbeurteilungsinstrumenten (etwa: Qualitätsmanagementsystemen, Stundenkalkulationssystemen oder Begleitgruppen) und damit zu einem Wandel der ‚Mertonwelt‘ beitrage, insofern nämlich die disziplinengebundenen resultatbezogenen Normen durch prozedurale Normen, die auf disziplinenunabhängigen Organisationen basieren, ergänzt, gar ersetzt würden. Pointiert: „Der Universalismus der Resultate wird ersetzt durch einen Universalismus der Firma.“

Den Abschluss dieser Einheit markierte – wiederum den Faden generalisierender Überlegungen zum Problemfeld transdisziplinärer Forschung aufnehmend – Christian Pohl (ETH Zürich) mit *„Besonderheiten der Evaluation transdisziplinärer Forschung“*. Unter Verweis auf verschiedene Charakterisierungsmöglichkeiten transdisziplinärer Forschung stellte er insbesondere vier Charakteristika heraus, die sich in bestimmten Evaluationsfragen niederschlagen:

- die gesellschaftsbezogene Problemidentifizierung und -strukturierung („Sind die involvierten Akteure bzw. Disziplinen relevant?“ „Sind die Forschungsfragen originell und zweckdienlich?“),
- die Einbettung in Lebenswelt und Wissenschaft („Wird der ‚State of the Art‘ in Wissenschaft und Lebenswelt erfüllt?“),
- die Integration („Sind die Formen der Zusammenarbeit zweckdienlich?“ „Sind die Mittel der Integration informiert ausgesucht?“)
- sowie das rekursive Forschungsdesign („Wird die transdisziplinäre Forschung als Mittel des gezielten Lernens eingesetzt?“).

3 Die Evaluation der Evaluation

Betrachtet man die weit gespannten Problemfelder, die sich im Kontext der Evaluationsforschung auftun, dann verwundert immer weniger der intensive wissenspolitische Diskurs um die Anwendungsbedingungen von Evaluationsinstrumenten zur Messung wissenschaftlicher

Qualität. Denn einerseits sind die Möglichkeiten der Feststellung von Qualität stark an je unterschiedliche disziplinspezifische Praktiken gebunden, andererseits müssen vor dem Hintergrund von Regulationsbemühungen sowie Anwendungs- und Transdisziplinaritätsbestrebungen die Generalisierungschancen von Evaluationsinstrumenten ausgelotet werden. Was sind also die Grenzen von Evaluation? Auch wenn Probleme der Vergleichbarkeit entstehen, so sind doch der zeitvariante Charakter von Evaluationen, das prozessbezogene Qualitätsmanagement und die kritische Indikatorenreflexion als allgemeine Anforderungen an den Einsatz von Evaluationsinstrumenten hervorzuheben. Evaluation kann dann zu einem vertretbaren wissenspolitischen Instrument werden, wenn es die Evaluation der Evaluation selbst wiederum auf Dauer stellt. Ansätze hierfür zeigen sich nicht nur in der Transdisziplinaritätsforschung, sondern auch bei der Schaffung von übergreifenden Wissenschaftsorganisationen wie dem IFQ.

Anmerkung

- 1) Das Referat bezog sich auf Ergebnisse des BMBF-Projekts ‚Auswirkungen der evaluationsbasierten Forschungsförderung an Universitäten auf die Inhalte der Forschung‘.

« »